

# Implications of Three Causal Models for the Measurement of Halo Error

Sebastiano A. Fisicaro, Wayne State University

Charles E. Lance, University of Georgia

The appropriateness of a traditional correlational measure of halo error (the difference between dimensional rating intercorrelations and dimensional true score intercorrelations) is reexamined in the context of three causal models of halo error. Mathematical derivations indicate that the traditional correlational measure typically will *underestimate* halo error in ratings and can suggest no halo error or even “negative” halo

error when positive halo error actually occurs. A corrected correlational measure is derived that avoids these problems, and the traditional and corrected measures are compared empirically. Results suggest that use of the traditional correlational measure of halo error be discontinued. *Index terms:* halo, halo effect, halo error, performance ratings, rating accuracy, rating errors.

Halo error has long been a concern to scientists and practitioners as a source of inaccuracy in interpersonal judgments and performance evaluations (Fisicaro, 1988). Typically, these judgments or evaluations are expressed as numerical ratings. According to Thorndike (1920), one consequence of halo error is to cause dimensional rating intercorrelations to be higher than they should be, thereby making otherwise conceptually distinguishable dimensions of behavior appear to be more highly related than they actually are.

As noted by Pulakos, Schmitt, and Ostroff (1986) in their critique of halo error measures, it cannot be concluded that halo error has occurred simply by examining the level of dimensional rating intercorrelations (observed halo)—that is, a baseline against which observed halo can be compared is necessary in order to infer the presence of halo error. An appropriate baseline index would indicate the expected level of dimensional rating intercorrelations when halo error is *not* being committed. Traditionally, researchers have considered dimensional intercorrelations of actual ratee behaviors or true scores (true halo) to represent such a baseline (Feldman, 1986). Thus, it is common for researchers to use the difference between dimensional intercorrelations of ratings and dimensional intercorrelations of actual ratee behaviors (observed halo minus true halo) as a measure of halo error (Fisicaro, 1988). In fact, Pulakos et al. (1986) argued that this difference measure is the most conceptually appropriate measure of halo error. This measure will be referred to as the *traditional* correlational measure of halo error.

This paper (1) reviews three well-known conceptual definitions of halo error; (2) shows that these definitions give rise to three causal models of halo error; (3) demonstrates, in the context of these models, that the traditional correlational measure of halo error is appropriate only under restrictive—and typically unrealistic—assumptions; (4) derives a more appropriate, corrected correlational measure of halo error based on the three causal models; and (5) compares the traditional and corrected correlational measures empirically.

---

**APPLIED PSYCHOLOGICAL MEASUREMENT**

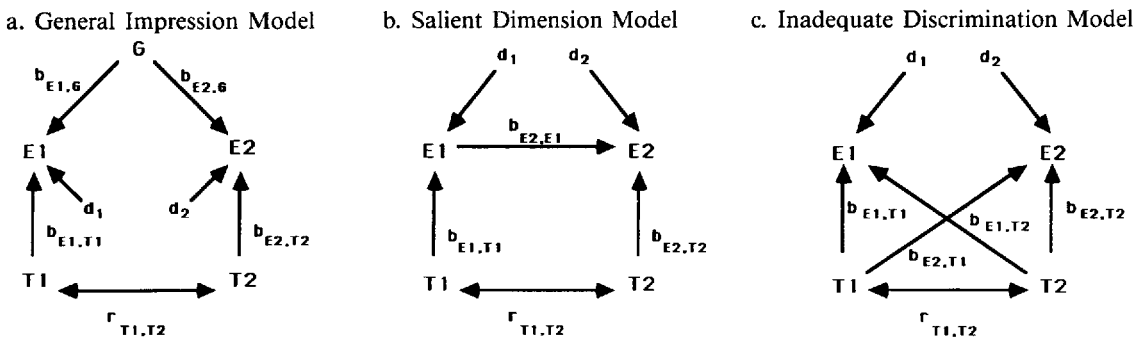
Vol. 14, No. 4, December 1990, pp. 419-429

© Copyright 1990 Applied Psychological Measurement Inc.  
0146-6216/90/040419-11\$1.80

### Three Conceptions of Halo Error

Three widely accepted conceptual definitions of halo error give rise to the three causal models of halo error shown in Figure 1. In these models single-headed arrows represent causal influences, double-headed arrows signify (exogenous) correlations ( $r$ s),  $G$  refers to a rater's general impression,  $T1$  and  $T2$  refer to actual ratee behavior on two different dimensions of behavior,  $E1$  and  $E2$  refer to rater evaluations of ratee behavior on the same two dimensions,  $b$ s represent path coefficients, and  $d$ s represent disturbance terms. For the sake of clarity, and without loss of generality, only two dimensions are included in these models.

Figure 1  
 Three Causal Models of Halo Error



As is common for algebraic convenience, all variables are assumed to be expressed in standard score form (e.g., James, Mulaik, & Brett, 1982). Also assumed are (1) linearity in variables and equations, (2) that  $d$ s are uncorrelated with explanatory variables and each other, and (3) that ratee behaviors generally are correlated across different dimensions (i.e.,  $r_{T1,T2} \neq 0$ ). Subsequently considered, however, is the special case in which ratee behaviors are uncorrelated across dimensions (i.e.,  $r_{T1,T2} = 0$ ).

#### General Impression Model

Halo error has been defined as “the tendency of a rater to allow overall impressions of an individual to influence the judgment of that person’s performance along several quasi-independent dimensions of job performance” (King, Hunter, & Schmidt, 1980, p. 507), or “the influence of a global evaluation on evaluations of individual attributes of a person” (Nisbett & Wilson, 1977, p. 250). The causal model implied by these definitions is shown in Figure 1a, a “General Impression” model. Here, a rater’s general impression ( $G$ ) is shown to have a common causal effect on dimensional evaluations  $E1$  and  $E2$  ( $b_{E1,G}$ ,  $b_{E2,G}$ ), which serves to “inflate” the correlation between  $E1$  and  $E2$  ( $r_{E1,E2}$ )—that is, according to this model, the source of halo error is the common causal effect of a rater’s general impression on dimensional evaluations (Landy, Vance, Barnes-Farrell, & Steele, 1980).

#### Salient Dimension Model

Halo error also has been defined conceptually as the influence of a rater’s evaluation of ratee behavior on one or more salient dimensions on evaluations of behavior on other dimensions (Anastasi, 1988; Blum & Naylor, 1968), or “the tendency for an evaluator to let the assessment of an individual

on one trait influence his or her evaluation of that person on other traits” (Robbins, 1989, p. 444). For example, a rater’s judgment of the quality of a ratee’s research record could influence the rater’s judgment of the ratee’s teaching effectiveness. This “Salient Dimension” model is shown in Figure 1b. Here, the source of halo error is a rater’s evaluation on a salient dimension (E1) directly influencing the rater’s evaluation on a second, less salient dimension (E2) (i.e.,  $b_{E2,E1}$  “inflates”  $r_{E1,E2}$ ).

When there are more than two dimensions, an evaluation on a salient dimension is a common cause of evaluations on other, nonsalient dimensions. For example, if a third, nonsalient dimension were included in Figure 1b, E1 (an evaluation on the salient dimension) would have a common causal effect on E2 and E3. In this case, sources of halo error would include (1) the direct effect of E1 on E2 ( $b_{E2,E1}$ ), (2) the direct effect of E1 on E3 ( $b_{E3,E1}$ ), and (3) the common causal effect of E1 on E2 and E3 ( $b_{E2,E1} b_{E3,E1}$ ). However, this additional complication does not alter any of the derivations or conclusions presented here.

### Inadequate Discrimination Model

Halo error also has been conceptualized as “a rater’s failure to discriminate among conceptually distinct and potentially independent aspects of a ratee’s behavior” (Saal, Downey, & Lahey, 1980, p. 415; see also DeCotiis, 1977; Murphy & Reynolds, 1988). An “Inadequate Discrimination” model of halo error suggested by this definition (see Figure 1c) attributes halo error to “cross effects” of ratee behaviors—that is, ratee behavior on one dimension influences evaluations of ratee behaviors on other dimensions (i.e.,  $b_{E1,T2}$  and  $b_{E2,T1}$  in Figure 1c), which, in turn, cause  $r_{E1,E2}$  to be “inflated.” These cross effects could result from a rater misinterpreting which ratee behaviors belong with which dimensions, due to factors such as inadequate rater familiarity with the dimensionality of target behaviors or from insufficiently concrete rating categories (Cooper, 1981).

### Comparison of the Models

One way to contrast the models in Figure 1 is to decompose the correlation between rater evaluations of ratee behavior on different dimensions ( $r_{E1,E2}$ ) into components that represent sources of covariation due to actual ratee behavior and halo error. For the General Impression, Salient Dimension, and Inadequate Discrimination models, these decompositions are, respectively,

$$r_{E1,E2} = b_{E1,T1} b_{E2,T2} r_{T1,T2} + [b_{E1,G} b_{E2,G}] \quad , \quad (1)$$

$$r_{E1,E2} = b_{E1,T1} b_{E2,T2} r_{T1,T2} + [b_{E2,E1}] \quad , \quad (2)$$

and

$$r_{E1,E2} = b_{E1,T1} b_{E2,T2} r_{T1,T2} + [b_{E1,T1} b_{E2,T1}] + [b_{E1,T2} b_{E2,T2}] + [b_{E1,T2} b_{E2,T1} r_{T1,T2}] \quad , \quad (3)$$

where bracketed terms indicate sources of halo error.

According to the three models shown in Figure 1, halo error results from (1) common causal effects of a rater’s general impression in the General Impression model (the bracketed term in Equation 1), (2) direct effects from evaluations on a salient dimension to evaluations on other dimensions in the Salient Dimension model (the bracketed term in Equation 2), and (3) common causal cross effects of ratee behavior in the Inadequate Discrimination model (the three bracketed terms in Equation 3). Thus, the three models are similar in the sense that each model represents a situation in which an overriding influence causes inflated dimensional intercorrelations of evaluations. The models differ, however, in terms of the locus and nature of the overriding influence.

### Reconsideration of the Traditional Correlational Measure of Halo Error

In the absence of halo rater error, Equations 1, 2, and 3 all reduce to

$$r_{E1,E2} = b_{E1,T1} b_{E2,T2} r_{T1,T2} \quad (4)$$

This situation is depicted in Figure 2. In this case, the path coefficients are correlation coefficients; thus,  $b_{E1,T1}$  and  $b_{E2,T2}$  become  $r_{E1,T1}$  and  $r_{E2,T2}$ , respectively, and Equation 4 becomes

$$r_{E1,E2} = r_{E1,T1} r_{E2,T2} r_{T1,T2} \quad (5)$$

The correlations,  $r_{E1,T1}$  and  $r_{E2,T2}$ , are termed *correlation accuracies* (Fiscaro, 1988; Sulsky & Balzer, 1988). In the absence of halo error, therefore, the dimensional intercorrelation of evaluations equals the correlation between actual rater behaviors on the two dimensions weighted by the product of the two correlation accuracy coefficients connecting actual rater behaviors and evaluations on the respective dimensions (i.e., an appropriate baseline index of  $r_{E1,E2}$  is  $r_{E1,T1} r_{E2,T2} r_{T1,T2}$ ). Therefore, the traditional correlational measure of halo error ( $r_{E1,E2} - r_{T1,T2}$ ) typically will *underestimate*—and perhaps seriously—the actual magnitude of halo error, because  $r_{T1,T2}$  is an appropriate baseline index for  $r_{E1,E2}$  only under conditions of *perfect* rater correlation accuracy (i.e.,  $r_{E1,T1} = r_{E2,T2} = 1.0$ ), which is unlikely.

An appropriate correlational measure of halo error, however, can be derived. Solving for halo error components in Equations 1 through 3 and substituting  $r_{E1,T1}$  and  $r_{E2,T2}$  for  $b_{E1,T1}$  and  $b_{E2,T2}$  produces, respectively,

$$r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2} = [b_{E1,G} b_{E2,G}] \quad (6)$$

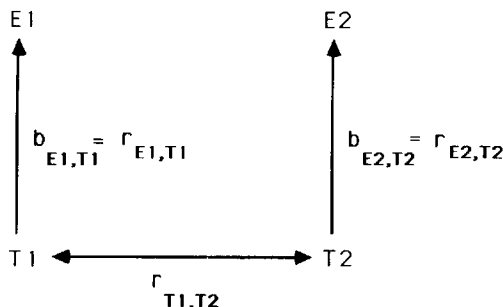
$$r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2} = [b_{E2,E1}] \quad (7)$$

and

$$r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2} = [b_{E1,T1} b_{E2,T1}] + [b_{E1,T2} b_{E2,T2}] + [b_{E1,T2} b_{E2,T1} r_{T1,T2}] \quad (8)$$

These results show that an appropriate measure of halo rating error for each of the models in Figure 1 is the difference  $r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2}$ , not the traditional correlational halo error measure ( $r_{E1,E2} - r_{T1,T2}$ ). In fact, if correlation accuracy is imperfect (i.e.,  $r_{E1,T1}$  and/or  $r_{E2,T2} < 1.0$ ), the traditional correlational measure can indicate (1) no halo error (i.e.,  $r_{E1,E2} - r_{T1,T2} = 0$ ) when halo error actually occurs (i.e.,  $r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2} > 0$ ); or (2) “negative” halo error (i.e.,  $r_{E1,E2} - r_{T1,T2} < 0$ ) when positive halo error actually occurs (i.e.,  $r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2} > 0$ , and  $r_{E1,T1} r_{E2,T2} < r_{E1,E2} / r_{T1,T2}$ ).

**Figure 2**  
 Model of Dimensional Rating Intercorrelation  
 in the Absence of Halo Error



This *corrected* correlational measure of halo error ( $r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2}$ ) requires estimates of (1) correlations between dimensional evaluations ( $r_{E1,E2}$ ), (2) dimension intercorrelations of actual ratee behaviors ( $r_{T1,T2}$ ), and (3) dimensional correlation accuracy ( $r_{E1,T1}$  and  $r_{E2,T2}$ ; see Fisicaro, 1988, Equation 8; Kozlowski, Kirsch, & Chao, 1986). However,  $r_{E1,T1}$  and  $r_{E2,T2}$  are easily estimated in studies in which dimensional rating intercorrelations and true score intercorrelations can be calculated. It is important to note that the corrected correlational measure of halo error appropriately indicates the extent to which halo rating error has occurred (i.e., the observed outcome of inflated dimensional rating intercorrelations), but does not distinguish among the possible underlying processes or rater errors that generated it (i.e., alternative models in Figure 1).

**Special Cases of Halo Error Models and Implications for the Measurement of Halo Error**

One special case of the General Impression (Figure 3a), Salient Dimension (Figure 3b), and Inadequate Discrimination (Figure 3c) models of halo error occurs if actual ratee behaviors are uncorrelated across dimensions (i.e.,  $r_{T1,T2} = 0$ ). Setting  $r_{T1,T2}$  equal to 0 in Equations 6, 7, and 8 produces, respectively,

$$r_{E1,E2} = [b_{E1,G} b_{E2,G}] \quad , \quad (9)$$

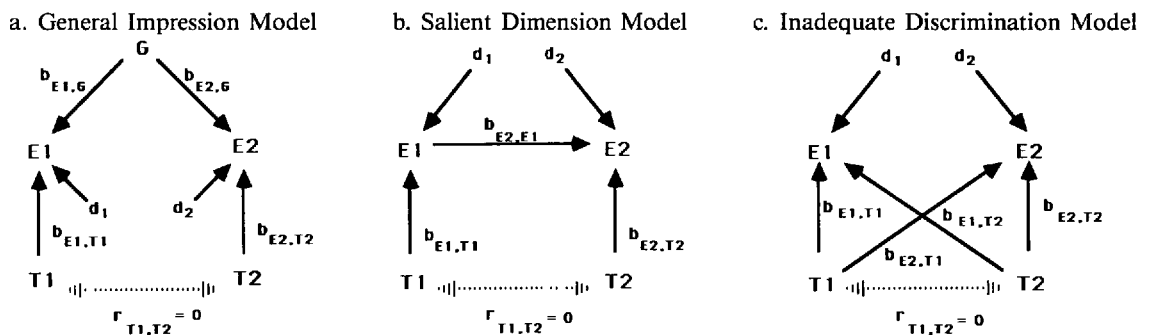
$$r_{E1,E2} = [b_{E2,E1}] \quad , \quad (10)$$

and

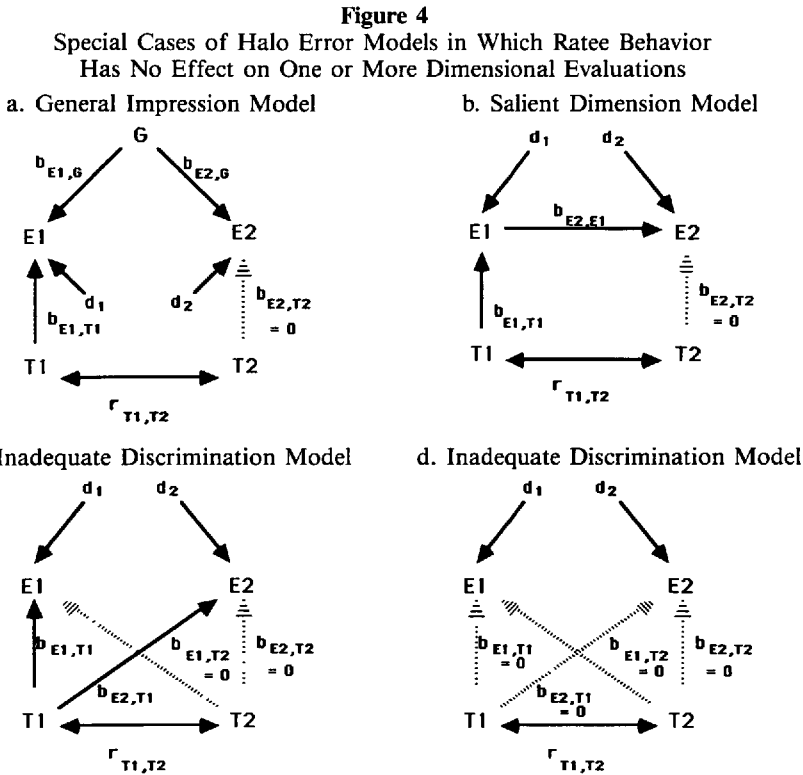
$$r_{E1,E2} = [b_{E1,T1} b_{E2,T1}] + [b_{E1,T2} b_{E2,T2}] \quad . \quad (11)$$

In this case, the traditional and corrected correlational halo error measures lead to identical conclusions, because *any* observed correlation among ratings ( $r_{E1,E2} \neq 0$ ) indicates the presence of halo error.

**Figure 3**  
 Special Cases of Halo Error Models With  
 Uncorrelated Dimensions of Behavior



A second special case of the models shown in Figure 1 occurs if actual ratee behavior on one or more dimensions has no influence on the rater's evaluation of the ratee on that dimension (i.e., correlation accuracy equals 0). For the General Impression model, it matters neither conceptually nor functionally whether  $r_{E1,T1} = 0$ , and/or  $r_{E2,T2} = 0$  (the case shown in Figure 4a). In either case, Equation 6 reduces to Equation 9, and any  $r_{E1,E2} \neq 0$  indicates the presence of halo error. Here, the corrected correlational measure ( $r_{E1,E2} - r_{E1,T1} r_{E2,T2} r_{T1,T2}$ ) produces an appropriate measure of halo error, because  $r_{E1,T1} r_{E2,T2} r_{T1,T2} = 0$ . However, the traditional correlational measure ( $r_{E1,E2} - r_{T1,T2}$ ) will *underestimate* halo



error, and will erroneously indicate “negative” halo error if, for example, both  $r_{T1,T2}$  and  $r_{E1,E2}$  are positive and  $r_{T1,T2}$  happens to be larger than  $r_{E1,E2}$ .

In the Salient Dimension model, it matters conceptually, but not functionally, whether  $r_{E1,T1}$  and/or  $r_{E2,T2} = 0$ . In either case, Equation 7 reduces to Equation 10. Figure 4b shows the more reasonable case in which ratee behavior has a significant impact on evaluations on a salient dimension (E1) but not on the other, nonsalient, dimension (E2). Here, for example, undergraduate students’ evaluations of an instructor’s teaching effectiveness could influence their evaluations of the instructor’s research productivity, something about which they may have little information. This situation is similar to that for the General Impression model in Figure 4a: Any  $r_{E1,E2} \neq 0$  indicates the presence of halo error (i.e.,  $r_{E1,E2} = b_{E2,E1}$ ), and the corrected correlational measure ( $r_{E1,E2} - r_{E1,T1}r_{E2,T2}r_{T1,T2}$ ) correctly indexes the magnitude of halo error present, because  $r_{E1,T1}r_{E2,T2}r_{T1,T2} = 0$ . However, the traditional correlational measure ( $r_{E1,E2} - r_{T1,T2}$ ) will again underestimate the extent of halo error, and will indicate negative halo error if  $r_{T1,T2}$  happens to be larger than  $r_{E1,E2}$ .

For the Inadequate Discrimination model, it matters both conceptually and functionally whether actual ratee behavior affects all, some, or none of the dimensional evaluations. Figure 4c shows the case in which actual ratee behavior on only one dimension influences evaluations (only T1 affects E1 and E2). Here,  $r_{E1,E2} = b_{E1,T1}b_{E2,T1}$  indicates rater halo error. The corrected correlational measure ( $r_{E1,E2} - r_{E1,T1}r_{E2,T2}r_{T1,T2}$ ) again indexes the magnitude of halo error correctly, because  $r_{E1,T1}r_{E2,T2}r_{T1,T2} = 0$ , and any  $r_{E1,E2} \neq 0$  indicates halo error. The traditional measure ( $r_{E1,E2} - r_{T1,T2}$ ), however, underestimates halo error.

In the extreme case that ratee behavior has no effect on any evaluations (Figure 4d), the corrected

correlational measure correctly indicates the absence of halo error, because  $r_{E_1, E_2} - r_{E_1, T_1} r_{E_2, T_2} r_{T_1, T_2} = r_{E_1, E_2} = 0$ . However, the traditional correlational measure ( $r_{E_1, E_2} - r_{T_1, T_2}$ ) will erroneously imply the existence of negative halo error as long as  $r_{T_1, T_2} > 0$ , and will erroneously imply the existence of positive halo error if  $r_{T_1, T_2} < 0$ . The former case, however, is more likely than the latter.

### An Empirical Comparison of Traditional and Corrected Correlational Halo Error Measures

#### Data Collection

The data used were collected by R. Tallarigo and the first author in 1986. Undergraduate psychology students ( $N = 52$ ) first completed a questionnaire measure of the perceived similarity of the dimensions they would subsequently use for rating. Students then were given a complete description of the rating task and dimensions. Next, the students viewed four videotaped lectures and rated each lecturer's performance on eight dimensions. Two lecturers spoke on the topic of self-fulfilling prophecy, and two others spoke on the effects of crowding on stress (see Murphy, Garcia, Kerkar, Martin, & Balzer, 1982, for a description of the videotapes and rating dimensions).

#### Measures

*Observed halo.* For each rater, an index of observed halo was calculated for each pair of rating dimensions as the correlation between ratings across ratees (i.e.,  $r_{E_1, E_2}$  in Equation 1). As is common and accepted practice (e.g., Fiscaro, 1988; Pulakos, 1984; Pulakos et al., 1986), an overall index of halo was calculated for each rater by first converting correlations using Fisher's  $r$ -to- $z$  transformation and then averaging the  $z$ s.

*True halo.* Correlations among actual ratee behaviors (true halo) were estimated from true score estimates (mean expert ratings) obtained by Murphy et al. (1982). The use of mean expert ratings as true score estimates is a relatively common practice in performance rating research. The justification for doing so derives from studies demonstrating the equivalence of mean expert ratings and more objective measurements (e.g., Smither, Barry, & Reilly, 1989). True score intercorrelations (e.g.,  $r_{T_1, T_2}$ ) were calculated for each pair of rating dimensions as the correlation between true scores across ratees. These correlations were transformed to  $z$ s, and the  $z$ s were then averaged to obtain the overall level of true halo. Because all raters viewed the same set of four videotaped lectures, true halo level was a constant (mean  $z = .473$ , back-transformed to  $r = .441$ ).

*Correlation accuracy.* For each rater, correlation accuracy scores were calculated for each of the eight rating dimensions as the correlation between dimensional ratings and ratee true scores. Again, these correlations were transformed to  $z$ s, and the  $z$ s were averaged.

*Traditional correlational measure of halo error.* For each rater, a traditional correlational measure of halo error was calculated for each rating dimension pair as the difference between observed halo and true halo for that pair of rating dimensions; these were transformed to  $z$ s and averaged.

*Corrected correlational measure of halo error.* For each rater, a corrected correlational measure of halo error was calculated for each rating dimension pair as the difference between observed halo (e.g.,  $r_{E_1, E_2}$ ) and true halo multiplied by raters' correlation accuracy scores for the rating dimensions in the pair (e.g.,  $r_{T_1, T_2} r_{E_1, T_1} r_{E_2, T_2}$ ); these, too, were transformed to  $z$ s and averaged.

#### Results

Table 1 contains descriptive statistics for the measures described earlier. Mean observed halo across raters was high (mean  $z = .85$ , back-transformed to  $r = .69$ ), but varied considerably across raters ( $-.231 \leq r \leq .992$ ). Mean correlation accuracy across raters (mean  $z = .367$ , back-transformed to

**Table 1**  
 Mean and Standard Deviation (SD) of Fisher's  
*r*-to-*z* Transformation, Minimum and Maximum *z*s, and  
 Back-Transformed Values of Mean, Minimum, and  
 Maximum *z*s (*r<sub>b</sub>*) for Five Measures

Measure	Mean	SD	Minimum	Maximum
Dimensional Rating Intercorrelation				
<i>z</i>	.850	.624	-.235	2.758
<i>r<sub>b</sub></i>	.691	-	-.231	.992
Dimensional True Score Intercorrelation				
<i>z</i>	.473	-	-	-
<i>r<sub>b</sub></i>	.441	-	-	-
Correlation Accuracy				
<i>z</i>	.367	.450	-.823	1.201
<i>r<sub>b</sub></i>	.353	-	-.677	.834
Traditional Correlational Halo Error Measure				
<i>z</i>	.112	.275	-.675	.589
<i>r<sub>b</sub></i>	.112	-	-.588	.529
Corrected Correlational Halo Error Measure				
<i>z</i>	.503	.388	-.137	1.691
<i>r<sub>b</sub></i>	.465	-	-.137	.934

*r* = .353) was somewhat lower than in earlier studies (e.g., Kozlowski & Kirsch, 1987) and also varied substantially across raters ( $-.677 \leq r \leq .834$ ).

There was a significant difference [paired sample  $t(51) = 6.20, p < .01$ ] in the level of halo error indicated by the traditional (mean  $z = .112$ , back-transformed to  $r = .112$ ) and corrected correlational measures (mean  $z = .503$ , back-transformed to  $r = .465$ ). This difference supports the idea that the traditional correlational measure can underestimate the magnitude of halo rating error actually present in ratings when correlation accuracy is imperfect.

Also examined was the frequency with which negative values were obtained for the traditional and corrected correlational halo error measures, and for observed halo. Negative values were obtained for 13 raters (25% of the sample) for the traditional correlational measure of halo error, for four raters (8%) for the corrected correlational measure, and for three raters (6%) for observed halo. Of the four raters who exhibited negative halo error according to the corrected correlational measure, two also exhibited negative observed halo.

These results suggest that the corrected correlational measure tends to identify more accurately the actual presence of negative halo error than does the traditional correlational measure. Furthermore, nine raters had negative values on the traditional correlational measure, but exhibited positive halo error on the corrected correlational measure and had positive observed halo. This supports the idea that the traditional correlational measure of halo error can indicate negative halo error when positive halo error has actually occurred.

### Discussion and Conclusions

These results support the arguments above that the traditional correlational measure of halo error can (1) underestimate the magnitude of halo error in the case of imperfect correlational accuracy, and (2) indicate negative halo error when positive halo actually occurs. In the context of the halo rater error models presented here, the traditional correlational measure is most appropriately viewed as a special case of the corrected correlational measure when correlation accuracy is perfect, which



is improbable at best; or when rater behaviors across dimensions are uncorrelated, which is also improbable, and in which case both the traditional and the corrected correlational halo error measures are equal simply to the observed dimensional rating intercorrelation. Thus, a more general and more appropriate measure for the three models of halo rater error presented here is the corrected correlational measure shown in Equations 6 through 8. This measure requires additional calculations of correlations between ratings and rater true scores (correlation accuracy), but these are routinely calculable in studies in which  $r_{r,T_r}$  (true halo) is estimable. Contrary to the position adopted by Pulakos et al. (1986), it is recommended that use of the traditional correlational measure of halo error be discontinued.

“Negative halo error” now has been reported in several studies (e.g., Fisticaro, 1988; Kozlowski & Kirsch, 1987; Murphy & Reynolds, 1988). These findings have been perplexing, because halo error is presumed to inflate—rather than attenuate—rating intercorrelations (Cooper, 1981). However, many findings of negative halo error may be attributable to use of the traditional correlational measure of halo error.

In the present study, the traditional correlational measure of halo error indicated negative halo error for 25% of the raters, whereas the corrected correlational measure indicated that positive halo error actually occurred in 69% of these cases. However, this is not to say that negative halo error cannot or does not occur. In fact, 8% of the raters in the present study exhibited negative halo error when the corrected correlational measure was used.

Furthermore, Lance, Fisticaro, and LaPointe (in press), and others (e.g., Murphy & Reynolds, 1988) have presented plausible rationales for the occurrence of negative halo error. The theoretical and empirical results reported here do not address findings of negative halo error reported or obtained with noncorrelational measures of halo error, such as variance measures (e.g., Fisticaro, 1988). Nonetheless, the traditional correlational measure of halo error, which fails to correct for correlation accuracy, likely has misled some researchers to infer the presence of negative halo error when positive halo error actually occurred along with imperfect correlation accuracy.

Halo error also appears to be a more pervasive phenomenon than has been acknowledged in the recent literature. There has been a shift in opinion away from traditional notions that ratings tend to be contaminated with systematic halo error toward the idea that halo error is “not nearly as serious a problem as has generally been presumed” (Kozlowski & Kirsch, 1987, p. 259). These conclusions, in large part, have been based on traditional correlational measures of halo error and, as has been demonstrated, the traditional correlational measure will tend to underestimate the prevalence and magnitude of positive halo.

In the context of the theoretical models presented here, the severity of this bias rests on the level of rater correlation accuracy: Low correlation accuracy levels lead to serious bias in the traditional correlational measure of halo error. The correlation accuracy level reported in Table 1 (mean back-transformed  $r = .353$ ) led to an estimate of halo error based on the traditional correlational measure that underestimated the actual magnitude of halo error by 76%. However, use of the traditional correlational measure of halo error, even with correlation accuracy levels in the .70s, can produce significant underestimates of the actual level of halo error.

Two limitations to the developments presented in this paper deserve mention. First, the corrected correlational measure is a measure of halo *rating* error. It does not distinguish among the General Impression, Salient Dimension, and Inadequate Discrimination models of halo error with respect to the type of *rater* error that may have occurred. The distinction between rater errors and rating errors is not new (e.g., Borman, 1975; Feldman, 1986; Vance, Winne, & Wright, 1983), but it seems to have been ignored in much of the recent literature. The developments presented here emphasize

the importance of maintaining this distinction, as well as suggesting that an appropriate correlational measure of halo rating error can be obtained irrespective of which of the three types of halo rater error may be operating.

Second, there may be relatively esoteric definitions of halo rater error that were not considered here, or definitions of halo rater error proposed in the future, for which the corrected correlational measure of halo error might not be appropriate. If so, an appropriate measure of halo rater error can be obtained by (1) calculating empirical estimates for path coefficients that represent the source of halo error (e.g.,  $b_{E1,G}$  and  $b_{E2,G}$  in the General Impression Model, Figure 1a), and (2) computing an estimate of the halo error component using these estimates (e.g.,  $b_{E1,G} b_{E2,G}$  for the General Impression Model and in Equation 1).

Finally, numerous studies of rater cognitive processes, rating errors, and context effects on ratings have been reported in recent years, yet little is known about the precise mechanisms that produce halo error in ratings. Identifying the causal mechanisms underlying halo rating error likely will lead to a better understanding of the situations and conditions conducive to halo rater error, and could have important implications for rater training. Thus, there should be increased emphasis on investigations of rater errors committed in evaluating others.

## References

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Blum, M. L., & Naylor, J. C. (1968). *Industrial psychology*. New York: Harper & Row.
- Borman, W. C. (1975). Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings. *Journal of Applied Psychology*, *60*, 556-560.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, *90*, 218-244.
- DeCotiis, T. A. (1977). An analysis of the external validity and applied relevance of three rating formats. *Organizational Behavior and Human Performance*, *19*, 247-266.
- Feldman, J. M. (1986). A note on the statistical correction of halo error. *Journal of Applied Psychology*, *71*, 173-176.
- Fisicaro, S. A. (1988). A reexamination of the relation between halo error and accuracy. *Journal of Applied Psychology*, *73*, 239-244.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data*. Beverly Hills CA: Sage Publications, Inc.
- King, L. M., Hunter, J. E., & Schmidt, F. L. (1980). Halo in a multidimensional forced-choice performance evaluation scale. *Journal of Applied Psychology*, *65*, 507-516.
- Kozlowski, S. W. J., & Kirsch, M. P. (1987). The systematic distortion hypothesis, halo, and accuracy: An individual-level analysis. *Journal of Applied Psychology*, *72*, 252-261.
- Kozlowski, S. W. J., Kirsch, M. P., & Chao, G. T. (1986). Job knowledge, rater familiarity, conceptual similarity and halo error: An exploration. *Journal of Applied Psychology*, *71*, 45-49.
- Lance, C. E., Fisicaro, S. A., & LaPointe, J. A. (in press). An examination of negative halo error in ratings. *Educational and Psychological Measurement*.
- Landy, F. J., Vance, R. J., Barnes-Farrell, J. L., & Steele, J. W. (1980). Statistical control of halo in performance ratings. *Journal of Applied Psychology*, *65*, 501-506.
- McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology*, *69*, 147-156.
- Murphy, K. R., Garcia, M., Kerker, S., Martin, C., & Balzer, W. K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology*, *67*, 320-325.
- Murphy, K. R., & Reynolds, D. H. (1988). Does true halo affect observed halo? *Journal of Applied Psychology*, *73*, 235-238.
- Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*, 250-256.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, *69*, 581-588.
- Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions within rater to measure halo. *Journal of Applied Psychology*, *71*, 29-32.
- Robbins, S. P. (1989). *Organizational behavior* (4th ed.).

Englewood Cliffs NJ: Prentice-Hall.

- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88, 413-428.
- Smither, J. W., Barry, S. R., & Reilly, R. R. (1989). An investigation of the validity of expert true score estimates in appraisal research. *Journal of Applied Psychology*, 74, 143-151.
- Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology*, 73, 497-506.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Vance, R. J., Winne, P. S., & Wright, E. S. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology*, 36, 609-620.

### Acknowledgments

*The second author was supported in part by Universal Energy Systems Contracts No. F49620-85-C-0013/SB5851-0360 and No. F49620-88-C-0053/SB5881-0378, Air Force Office of Scientific Services, Bolling AFB, Washington DC. Portions of this paper were presented at the 4th Annual Conference of the Society for Industrial and Organizational Psychology, Boston MA, U.S.A., April, 1989. The authors thank Karl Kuhnert, Bob Vandenberg, and two anonymous reviewers for their helpful comments on an earlier version of this paper.*

### Author's Address

Send requests for reprints or further information to Sebastiano A. Fisicaro, Wayne State University, Department of Psychology, 71 W. Warren Avenue, Detroit MI 48202, U.S.A.